The Path Forward for Successful Collaborative Research

Deb Agarwal, Mary Thompson
Distributed Systems Department

Abstract: *One of the central issues facing us as computer scientists is how to get collaboratory software used by the scientists for whom it was designed. The collaboratory program has succeeded in linking computer scientists with other scientists and providing software for use within these collaboratories. However, there remains the next step of moving this software into regular production use. This paper will explore some of the past experience and suggest some possibilities for the future.*

Over the past ten years, the DOE collaboratory development projects have explored and built a variety of scientific collaboration capabilities. These collaboratory projects have focused on a few main areas: remote instrument access, shared access to computing and data, and direct interaction between collaborators. Many of the collaboratory development efforts have been funded in partnership with a specific scientific user community with the scientific community committing to helping formulate requirements and using the tools during the course of the project. The idea has been for the requirements of the scientific community to directly drive the development efforts and thus improve the chances of adoption. Since user desire for, or at least acceptance, of collaboration is essential, this user-centered strategy is crucial for the long-term adoption of collaborative tools. However, this approach requires two directly competing goals: leading edge collaboratory development and release of robust, full-featured user tools.

Although this approach has served as an excellent means of gathering requirements and experience in the use of these tools by scientists in the course of their research, it has often had the negative effect of alienating users due to lack of robustness and support for stable versions of the tools. Those projects that have managed to maintain their user community have only done so by working 80-hour weeks (40 hours for development and 40 hours for support of users). Even so, these projects have still needed to push their software out to users before it is ready.

The pilot projects have identified many underlying functionalities that are important to effective collaboration. Development of middleware components to provide these functions has sometimes been funded as separate projects. In the best cases these projects have learned from the requirements and experiences of the pilots and have been able to return improved versions of the middleware to real users. Some examples of these projects include the Access Grid, the Globus Toolkit, electronic notebooks, customizable chat and portal software, authentication and authorization solutions, and group communication development. While usable middleware and tools can be deployed in the 3-5 year funding cycles, the hardening of these components and widespread adoption will take considerably longer.

It has become increasingly clear to those of us involved in the production of middleware and tools, that we need to work on three tiers: doing the significant computer science and sociological research to develop new functionalities; adapting current prototypes to the immediate needs of other scientists; and continuing to maintain and harden middleware and tools that are currently in use.

One of the goals of this NC program meeting should be to discuss policies and mechanisms to provide research projects, pilot deployment efforts, and the hardening and support for those tools that reach wide deployment and use. Some of the obvious questions that must be answered at this meeting include:

- What research projects need to be pursued?
- What existing tools are researchers relying on that must be maintained to prevent gradual decay of the infrastructure on which the researchers depend?
- How can we provide the long-term support within the NC program, or some other DOE program?

Experience has shown that most of the software that is designed to meet specific needs of DOE scientists will not be picked up from the prototype level and turned into products by industry. This is especially true of middleware, which is considered a commodity that should be freely available. There are several reasons why even application software has not been commercialized. First the needs and working environment of scientists are often different from those of the business and consumer communities, so the market is very small. Second, scientists usually want the ability to change and customize software to meet their changing needs, so they are much happier with open source software that permits this. Third, scientists do not want to spend a lot of money on software; they are used to writing and sharing software in an open source model. At best some of the ideas from our prototypes may be incorporated into commercial products that will be targeted to somewhat similar uses.

The current mode of scientists writing and sharing code informally, could be expanded and maybe standardized across science domains including computer scientists. If we envisage our collaboratory software being distributed through something like SourceForge, we could expect that the maintenance and further development of the third tier tools might gradually pass to the users of the tools. However, it is likely that a model like NMI or ACTS would be needed where there is some centralized support of the tools.

This model promises to be successful only if there is some explicit funding support. Computer science groups would need to be funded to package and install their code in a distribution framework and to provide the initial support. Other scientists would need to be funded as they took over the support of the tools that they rely on. Computer science groups would also need to be funded to do base research to build the next generation of tools. All three components of this model are essential to long-term success.

Acknowledgements: *Doug Olsen, Dan Gunter, Brian Tierney and other members of DSD have contributed to the discussion of this issue.*